



Apr 08, 2025

RNAseq libraries preparation, analysis and data processing

 [Nature Cancer](#)

DOI

<https://dx.doi.org/10.17504/protocols.io.j8nlkd486g5r/v1>

Laura Antonucci^{1,2}, Isidoro Cobo^{3,2}, Chiara Nicoletti^{4,5}, Angeles Duran Molina⁶, Michael Karin^{1,2}

¹Laboratory of Gene Regulation and Signal Transduction, Departments of Pharmacology and Pathology, University of California San Diego School of Medicine;

²La Jolla, CA 92093, USA; ³Department of Cellular & Molecular Medicine, UCSD School of Medicine;

⁴Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute;

⁵La Jolla, CA 92037, USA;

⁶Department of Pathology and Laboratory Medicine and Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA



Laura Antonucci

University of California, San Diego

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.j8nlkd486g5r/v1>



External link: <https://doi.org/10.1038/s43018-025-01003-3>

Protocol Citation: Laura Antonucci, Isidoro Cobo, Chiara Nicoletti, Angeles Duran Molina, Michael Karin 2025. RNAseq libraries preparation, analysis and data processing. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.j8nlkd486g5r/v1>

Manuscript citation:

Antonucci, L., Li, N., Duran, A. *et al.* Self-amplifying NRF2–EZH2 epigenetic loop converts KRAS-initiated progenitors to invasive pancreatic cancer. *Nat Cancer* **6**, 1263–1282 (2025). <https://doi.org/10.1038/s43018-025-01003-3>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: April 08, 2025

Last Modified: April 08, 2025

Protocol Integer ID: 126325

Keywords: rnaseq libraries preparation, processing rnaseq library, rnaseq library, transcription factor enrichment analysis, mm10 genome, rna extraction, gene expression, genome, mouse pancreata, murine cell line, gene, rna

Abstract

RNAseq libraries were prepared from mouse pancreata and murine cell lines (CTRL, EZH2KO, EEDKO UN-KC6141) following standard protocols. After RNA extraction and poly-A enrichment, libraries were generated and sequenced on HiSeq 4000 or NextSeq 500 platforms. Reads were aligned to the mm10 genome, and gene expression was quantified and analyzed for differential expression using HOMER, Cufflinks, or DESeq2. Pathway and transcription factor enrichment analyses were performed with GSEA.

1 RNAseq libraries preparation for CAE experiments in WT and *Kras*^{G12D/PEC} pancreata, and for CTRL, EZH2^{KO}, and EED^{KO} UN-KC6141 was performed as described elsewhere(1).

RNA and DNase treatment was carried out using Direct-zol RNA MicroPrep kit (Zymoresearch, #11-33MB).

1µg total RNA was enriched in poly-A tailed RNA transcripts by double incubation with Oligo d(T) Magnetic Beads (NEB, S1419S) and fragmented for 9 min at 94°C in 2X Superscript III first-strand buffer containing 10mM DTT (Invitrogen, #P2325).

Reverse-transcription (RT) reaction was performed at 25°C for 10 min followed by 50°C for 50 min.

RT product was purified with RNAClean XP (Beckman Coulter, #A63987).

Libraries were ligated with dual UDI (IDT) or single (Bioo Scientific), PCR-amplified for 11-13 cycles, and size-selected using one-sided 0.8× AMPure cleanup beads, quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and sequenced on a HiSeq 4000 or NextSeq 500 (Illumina).

Analysis was performed as indicated. FASTQ sequencing files were mapped to the mm10 reference genomes using STAR with default parameters. Biological and technical replicates were used in all experiments. The quantification of transcripts was performed using analyzeRepeats.pl (HOMER) with parameters -condenseGenes -count exons -noadj. Principal Component Analysis (PCA) was obtained based on the Transcripts Per kilobase Million (TPM) on all genes of all samples. The expression value for each transcript was calculated using the analyzeRepeats.pl tool of HOMER with the following parameters:condenseGenes -count exons -tpm. Differential expression analysis was calculated using getDiffExpression.pl tool of HOMER using default parameters (FDR <0.05 and log2fold change > 1 or < -1). Pathway analyses were performed using the Molecular Signature Database of GSEA(2, 3).

RNAseq preparation and analysis for WT, *Kras*^{G12D/PEC}, *Kras*^{G12D/PEC}; *Nrf2*^{Act-PEC}, and *Nrf2*^{Act-PEC} pancreata, and for CTRL and EZH2^{KO} UN-KC6141 cells was performed as follows. Single-end 50 bp reads were obtained by RNA sequencing (RNAseq). The FASTQC module was run on FASTq files to check data quality. Quality scores for raw reads were Sanger transformed using FASTq Groomer. FASTq Groomer outputs were aligned to mm10 genome using TopHat (-first strand) in local sensitive model. Aligned reads were sorted by coordinates using Sort BAM module. Gene expression estimates were calculated using Cufflinks using reference mm10 GTF file from iGenomes. Differential gene expression was calculated for all pairs using the CuffDiff module. For gene set enrichment analysis (GSEA), gene expression matrices were pooled from gene expression estimates from Cufflinks output and processed with human-translated gene symbols with 1,000 permutations using a t-test metric for gene ranking. Enrichment was tested using default v5.2 MSigDB gene sets. After mapping to human-translated gene symbols in GSEA, enrichment of transcription

factor target (TFT) binding motifs (c3 TFT MSigDB gene set, v7.0, n=610 gene sets) was performed using 1000 permutations and the t-test metric for gene ranking. TFT gene sets were filtered for FDR p value < .25 and sorted by NES scores. The top most negatively enriched gene sets mapping to known TF are depicted.

RNAseq data analysis for CTRL, EED^{KO}, and EZH2^{KO} UN-KC6141 cells was performed as follows.

Data were checked for quality with FASTQC (v0.11.9) and aligned to the mm10 version of the mouse genome (mm10_UCSC_GRCm38.p3) with STAR (v2.7.3a). Gene expression quantification was achieved within the STAR alignment step, with the parameter `--quantMode GeneCounts`. Differential expression analysis was carried out with DESeq2 (v1.34.0), within the R environment (v4.1.3), selecting as differential genes with q-value < 0.05 and log2FC > |1|. Gene Ontology analysis was performed with GSEA PreRanked on gene sets from the Hallmarks7.5.1 repository, using the full list of genes tested in

the differential analysis step as input. GSEA plots were created using custom code, and Venn diagrams were made with Venny (v2.0.2).

Protocol references

1. J. S. Seidman *et al.*, Niche-Specific Reprogramming of Epigenetic Landscapes Drives Myeloid Cell Diversity in Nonalcoholic Steatohepatitis. *Immunity* **52**, 1057-1074 e1057 (2020).
2. V. K. Mootha *et al.*, PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
3. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545-15550 (2005).